

# Using Anchoring Vignettes to Assess Cross Cultural Comparability of CAHPS Ratings: A Pilot Study

Patricia M. Gallagher, Ph.D., F.J. (Jack) Fowler, Ph.D.  
Center for Survey Research, UMass Boston

&

Paul D. Cleary, Ph.D.  
Department of Health Care Policy, Harvard

# Acknowledgements



- **Matthew Cioffi, MS, Harvard**
- **Alison Hauser, MA, Center for Survey Research**
- **Vickie Stringfellow, BA, Center for Survey Research**
- **MassHealth Administrators and Members**

# Background



- **Evidence that CAHPS ratings vary by racial/ethnic group**
- **Issue: Not clear if this represents:**
  - Differences in health care experiences
  - Differences in the use of response options
  - Differences in expectations

# Overview of the Anchoring Vignettes Technique (1 of 2)



- Model proposed by Gary King, Christopher Murray, et al. (2004).
  - Respondents are asked to assess the care presented in hypothetical vignettes.
  - They are also asked to use the same response choices to self-assess the domain of care described in the vignettes.
  - Responses to the vignette items can be used to create comparable inter-rater measurements to adjust self-assessments

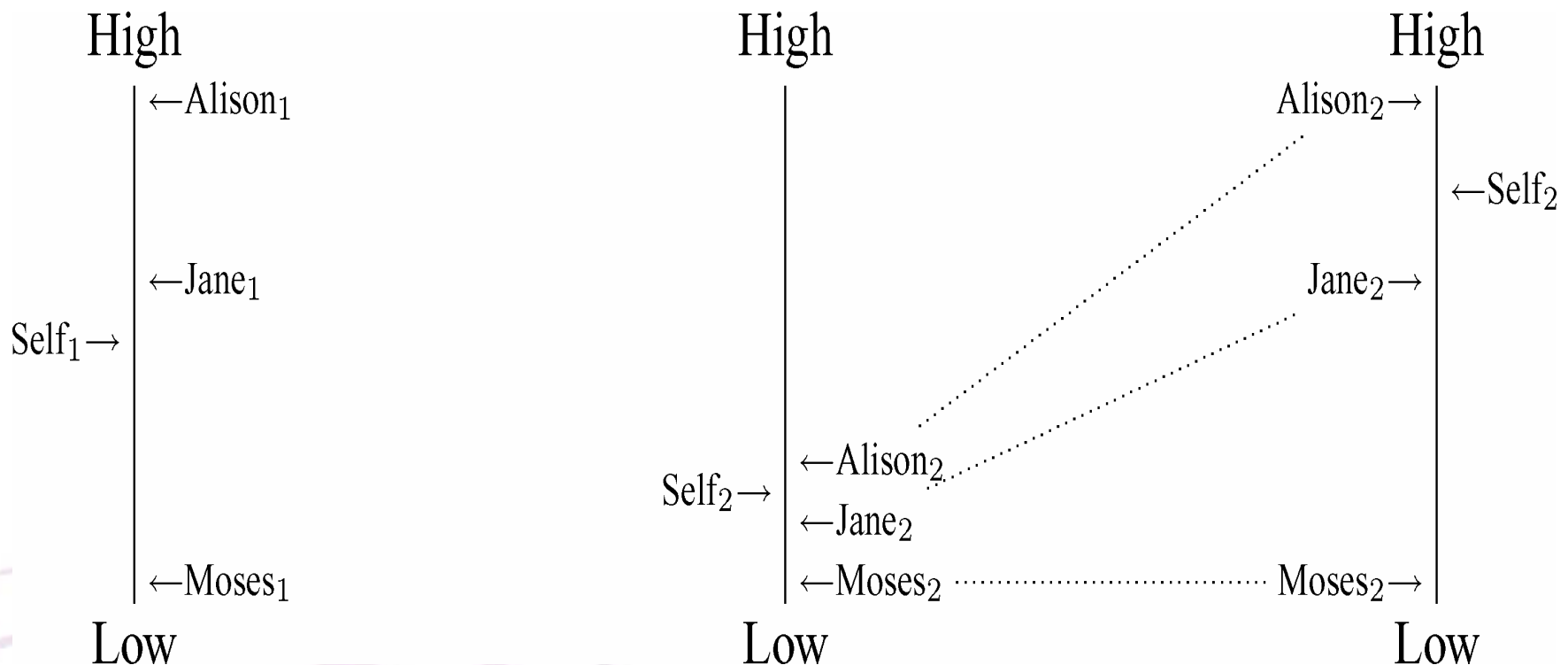
# Overview of the Anchoring Vignettes Technique (2 of 2)



- Responses to the vignettes are used as anchors for the self-assessment question
- Analysis involves specifying a joint CHOPIT (compound hierarchical ordinal probit) model for both the self-assessment item and the vignettes.

# Anchoring Vignettes Schematic

(King, Murray, et al. 2004, by permission)





# Pilot Test Methods



- Focus groups to develop and test vignettes (n=6).
- Translation of finalized vignettes to Spanish.
- Draw parallel samples of Latino, African American, and Non-Hispanic White adult MassHealth members (n=1400).
- Field CAHPS Core items and vignettes in a dual language instrument [Spring 2004].
- Mail survey: Advance Letter; ML1; PC; ML2

# Vignette #1



**When Mary goes to her personal doctor, she sometimes feels rushed during her time in the exam room. While the doctor explains some things carefully, sometimes Mary still has questions that have not been answered when she leaves the office.**

**We want to know your rating of Mary's personal doctor. Using any number from 0 to 10 where 0 is the worst personal doctor possible and 10 is the best personal doctor possible, what number would you use to rate Mary's personal doctor?**



# Vignette #2

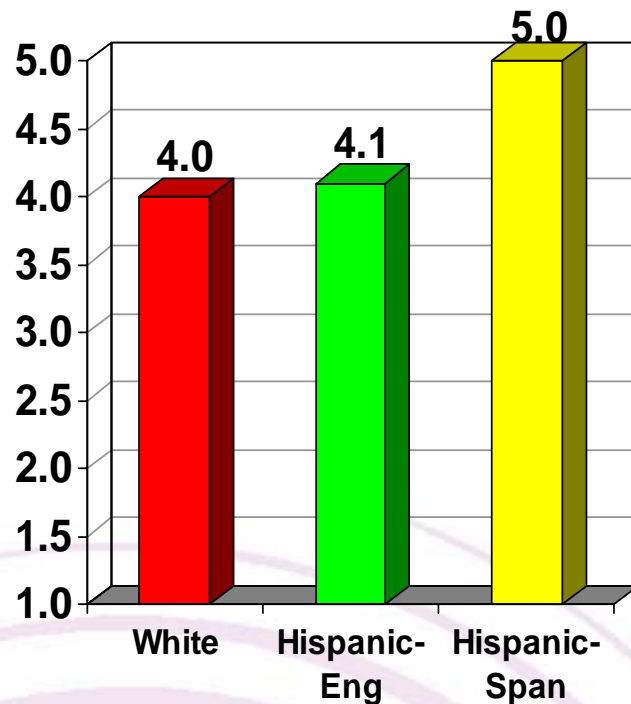


**Joe has to wait at least 2 months for an appointment with his personal doctor for check-ups. During the last year, his doctor treated him for heartburn for a month, but it turned out to be an ulcer.**

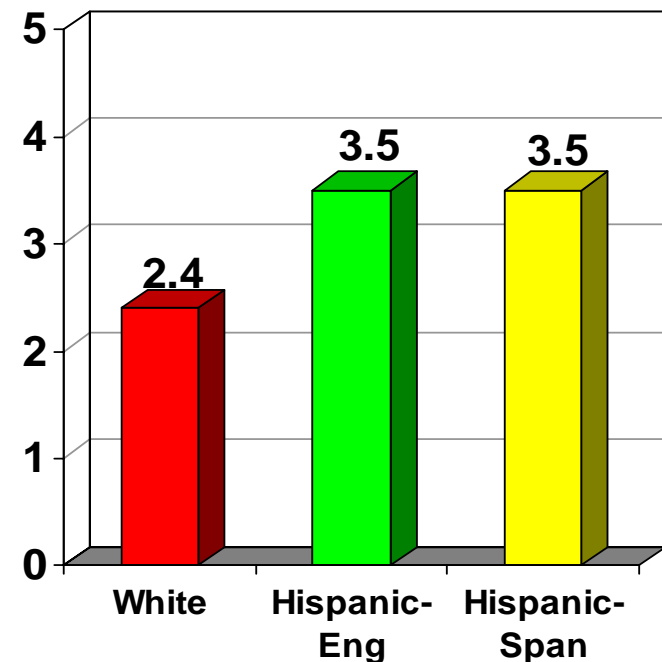
**Using any number from 0 to 10 where 0 is the worst personal doctor possible and 10 is the best personal doctor possible, what number would you use to rate Joe's personal doctor?**

# Mean VIGNETTE Ratings by Group: Adjusted for Health Status and Age

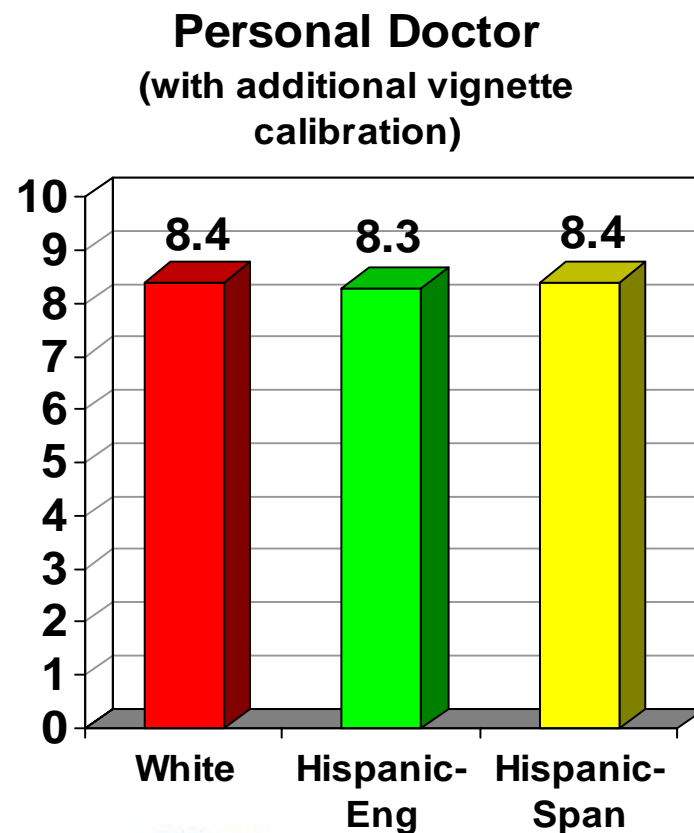
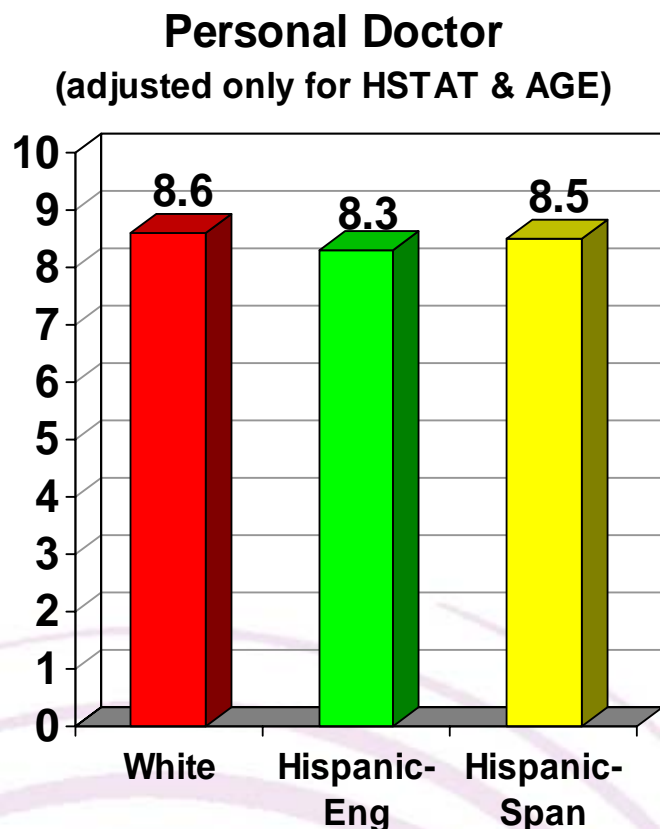
Vignette-Mary



Vignette-Joe



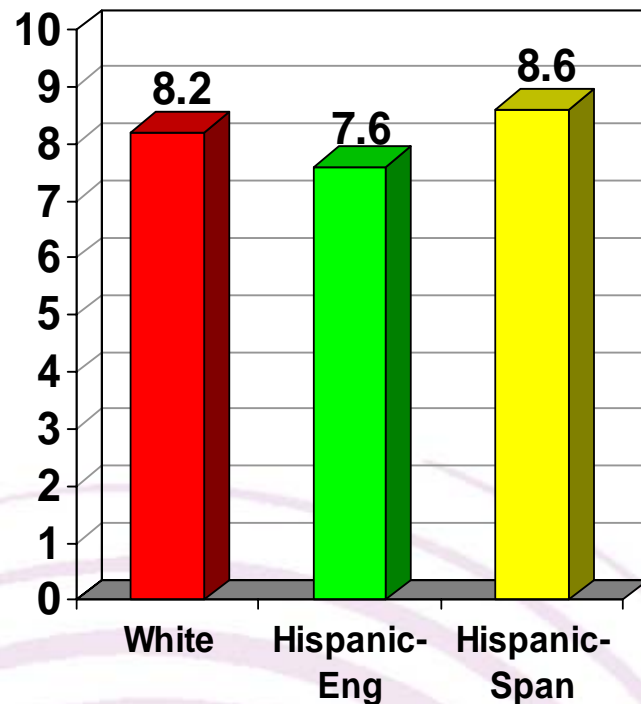
# Mean PERSONAL DOCTOR Ratings: HStat & Age Adjusted vs Vignette Adjusted



# Mean ALL CARE Ratings by Group: HStat & Age Adjusted vs Vignette Adjusted

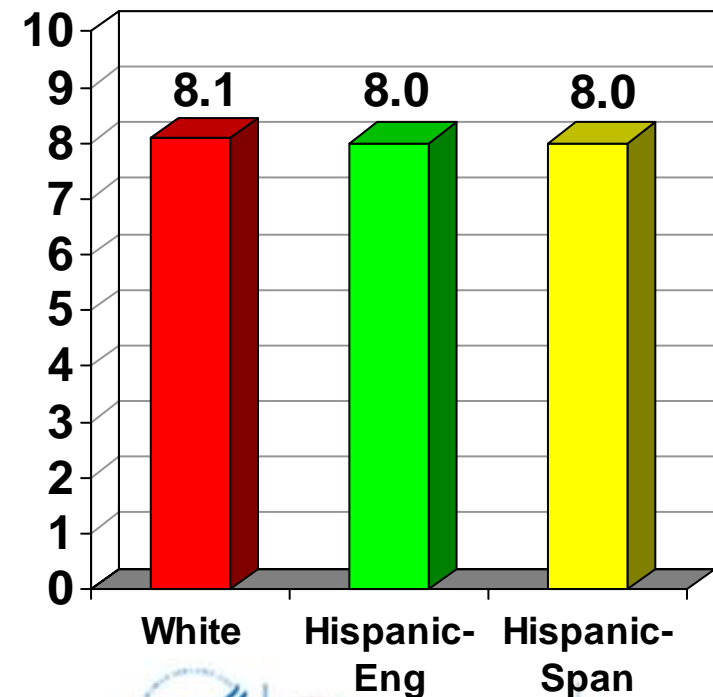
## All Care

(adjusted only for HSTAT & AGE)



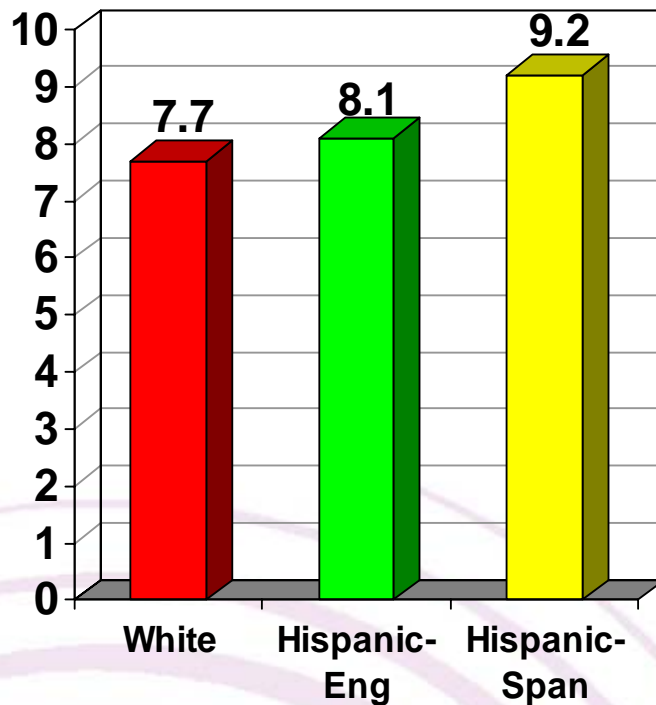
## All Care

(with additional vignette calibration)

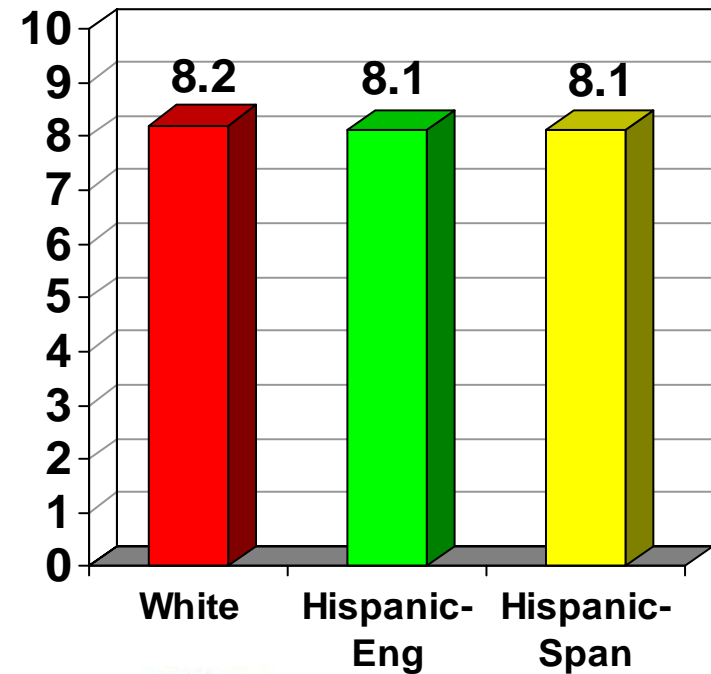


# Mean HEALTH PLAN Ratings: HStat & Age Adjusted vs Vignette Adjusted

**Health Plan**  
(adjusted only for HSTAT & AGE)



**Health Plan**  
(with additional vignette  
calibration)



# Conclusions



- **Latinos, particularly Spanish-speaking Latinos, appear to use ratings somewhat differently than White respondents.**
- **Consistent with previous findings, we have evidence [not presented owing to time limitations] that in reporting experiences there is less of an effect than in the reporting of ratings.**
- **More research is planned to build on the findings from these intriguing pilot results.**



# References



- King, Gary; Christopher J.L. Murray; Joshua A. Salomon; and Ajay Tandon. "Enhancing the Validity and Cross-cultural Comparability of Measurement in Survey Research," *American Political Science Review*, Vol. 97, No. 4 (December, 2003), 567-584; reprinted, with printing errors corrected, Vol. 98, No. 1 (February, 2004): 567-583.
- Rabef-Hesketh, S., A. Skrondal. Estimating chopit models in gllamm: Political efficacy example from King et al. (2002)
- Websites for additional information on Anchoring Vignettes:  
  
<http://GKing.Harvard.edu/vign/>  
  
[www.iop.kcl.ac.uk/loP/Departments/BioComp/programs/gllamm.html](http://www.iop.kcl.ac.uk/loP/Departments/BioComp/programs/gllamm.html)

# Questions?



100 Morrissey Boulevard  
Boston, MA 02125-3393  
P: 617-287-7200  
F: 617-287-7210  
[www.csr.umb.edu](http://www.csr.umb.edu)